AI Chains: Transparent and Controllable Human-AI Interaction by Chaining Large Language Model Prompts

Abstract

Although large language models (LLMs) have demonstrated impressive potential on simple tasks, their breadth of scope, lack of transparency, and insufficient controllability can make them less effective when assisting humans on more complex tasks. In response, we introduce the concept of Chaining LLM steps together, where the output of one step becomes the input for the next, thus aggregating the gains per step. We first define a set of LLM primitive operations useful for Chain construction, then present an interactive system where users can modify these Chains, along with their intermediate results, in a modular way. In a 20-person user study, we found that Chaining not only improved the quality of task outcomes, but also significantly enhanced system transparency, controllability, and sense of collaboration. Additionally, we saw that users developed new ways of interacting with LLMs through Chains: they leveraged sub-tasks to calibrate model expectations, compared and contrasted alternative strategies by observing parallel downstream effects, and debugged unexpected model outputs by "unit-testing" sub-components of a Chain. In two case studies, we further explore how LLM Chains may be used in future applications.